Tutors perception and modelling of the learner

Work package	Corpus of affective learner's expressions and interaction data (WP4)	
Task	Task 4.1	
Dissemination Level	Public	
Publishing date	Contractual M18	Actual
Deliverable	D4.1	
WP / Task responsible	UoB	
Contact person	Ginevra Castellano, UoB	
Contributors	UOB, JacobsUni, INESC-ID, UGOT	
Short abstract	This report presents the results of Task 4.1. It describes the requirements for the development of the robotic tutor's perception capabilities, with a particular focus on automatic affect recognition. Additionally, it presents the design and development of two studies performed in order to collect representative affective learner's expressions and interaction data.	
Keywords	Engagement, learning, artificial agents, affect recognition	
Documents	Deliverable 4.1	

Contents

1. Introduction

Work Package (WP) 4 focuses on the development of a computational framework for the tutor's perception of the learner and learner models that integrates behavioural and contextual information. Specifically, WP4 aims to design, develop and evaluate a system for the automatic recognition of learner affect. In order to develop such a system, representative data is required for training purposes. For this reason, a number of experiments that focused on collecting user affective expressions (user engagement, i.e., task and social engagement) in human-robot interactions supported by a multi-touch table have been designed, developed and conducted in Task 4.1 during the first 18 months of the project. Specifically, a pilot study was performed to inform the design of a corpus collection via Wizard-of-Oz (WoZ) of the tutor's interactions with students in a classroom for training the affect recognition system to be developed in Task 4.3. Following the pilot study, a WoZ study was performed in a classroom environment to collect a corpus of learner affective expressions in the context of the geography map reading scenario developed in WP2.

In summary, the objectives of Task 4.1 are:

- Corpus collection via WoZ study of the tutor's interaction with students in a classroom for training the affect recognition system developed in Task 4.3
- Video recordings of learners' behavioural expressions and synchronised contextual information related to the learning task and the tutor's behaviour
- Continuous annotation of learners' affective states

This report presents the results of Task 4.1, describing the requirements for the development of the robotic tutor's perception capabilities, with a particular focus on automatic affect recognition. Additionally, it presents the design and development of two studies performed in order to collect representative affective learner's expressions and interaction data.

The report is organised as follows. Section 2 discusses motivation behind the need to endow the robotic tutor with automatic engagement detection abilities, as well as related work in the area. Section 3 provides an overview of the requirements of the tutor's perception capabilities. Section 4 and 5 present a pilot study on user engagement and a WoZ study through which we collected a corpus of learner's expressions during the interaction with the robotic tutor. Preliminary data analysis from these studies is also presented. Finally, Section 6 discusses conclusions and future work in relation to Task 4.2 and Task 4.3.

2. Automatic engagement recognition: motivation and related work

During social interactions we continually output social cues containing indicators of our internal affective state. These cues are often displayed sub-consciously through our facial expressions, posture and hand gestures (Gunes & Schuller, 2012) and will often be extremely subtle. It has been shown that these cues are capable of establishing, maintaining and signifying our present level of engagement with an interaction (Glowinski & Mancini, 2011; Shic et al., 2008). However, the quantity, quality and depth of cues being emitted during human interactions are vast and extremely complex to decipher using current state-of-the-art technology.

In human-robot interaction scenarios, establishing an engaging interaction is a two-fold affair. Firstly the human must be able to perceive the robot as a social partner capable of an engaging interaction, and secondly the robot must be able to recognise and maintain the human interactant's level of engagement. Here, we are interested in two primary types of engagement: 1) *Social engagement*, defined as "the process by which two (or more) participants establish, maintain and end their perceived connection during interactions which they jointly undertake" (Sidner et al., 2004), and "the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and continuing the interaction" (Poggi, 2007), and 2) *Task engagement*, which shares common attributes of flow experience, characterised by elements of attention, concentration and enjoyment with a task (Shernoff et al., 2003).

The phenomenon of engagement is a widely discussed topic in the field of human-robot interaction (HRI), but in our work we wish to advance the current state-of-the-art by developing an on-line automatic engagement detector. In related work, Rich and collaborators (2010), explored the issue of recognising engagement in detail, modelling an interaction based on four different connection events involving gesture, speech, directed gaze, mutual gaze, conversational adjacency pairs and backchannels. Furthermore, Castellano and collaborators (2012), used game and social content based features to successfully predict engagement during child-robot interactions using trained Support Vector Machine (SVM) based models. Other projects have shown that prediction and anticipation is somewhat symbiotic to intention recognition, an important concept in social robotics (Sakita et al., 2004; Awais & Henrich, 2010; Jung et al., 2010; Shindev et al., 2012).

3. Tutor's perception capabilities: requirements and sensors

Automatic affect recognition requires the perception of several different modalities connected with each other in order to provide accurate recognition in real-time. External stimuli can be perceived and analysed through a variety of sensors that transform the signal into a digital representation readable from the system. Detecting engagement in real-time environments can be challenging and requires sensors to recognise various modalities of expression from users such as their body movement, facial features, physiological signals, etc. In the EMOTE project we evaluated a number of sensors for the perception of user behaviours in order to find the most efficient in school environments without complicating the overall architecture or compromising the performance of the system for real-time operations.

Our first basic requirement for the system was to recognise emotional facial expressions in real-time. Most of the facial recognition software development kits (SDKs) support only detection of basic emotions, while a few of them support detection of facial action units (AUs) in real-time. We conducted an evaluation of several sensing platforms for the development of the tutor perception modules. Facial expression recognition SDKs were compared: these included third-party software such as OKAO[1], Noldus FaceReader[2], CERT [3]and Kanako[4]. FaceReader performed quite well in our tests, however, only automatic detection of basic emotions is possible. The Kanako SDK is currently under development and the hardware requirements made it difficult to work on the different hardware that each consortium partner has. CERT SDK comes free for academic use and recognises all AUs however, the academic version does not support real-time recognition. OKAO SDK does not offer any individual action units but is bundled with many features such as expression recognition, eye gaze information, smile estimation etc. Additionally, OKAO is free of charge for academic usage.

For that reason we decided to use the OKAO SDK. Since it does not support the automatic detection of AUs we utilised the real-time face recognition of the Microsoft Kinect[5]sensor. The Kinect sensor provides the recognition of several AUs along with multiple head position and orientation detection and some basic skeleton tracking. At this stage of the project we decided to use the Kinect sensor for gaze detection as the results provided sufficient performance for initial gaze estimation. However, since OKAO already supports eye gaze estimation, the information of these two sensors will be combined in order to provide a more accurate gaze

[1] OKAO Vision SDK, www.omron.com/r_d/coretech/vision/okao.html
[2] FaceReader API, www.noldus.com/facereader/facereader-api
[3] CERT SDK, www.emotient.com/cert
[4] Kanako SDK, University of Nottingham
[5] Microsoft Kinect Sensor for Windows, www.microsoft.com/en-us/kinectforwindows

estimation based on head and eye orientation. The skeleton tracking feature is extracted for estimating the location and height of the user's head in order for the robot to gaze at the face of the user. Finally, the sensor also provides information regarding the depth of view therefore, predicting the average distance of the user from the sensor.

Lastly, we decided to utilise an electronic wrist sensor for capturing users' electro-dermal activity during their interactions with the system. The sensor we used is the Q Sensor from Affectiva[6]. This sensor can read users' galvanic skin response through the skin and transmit it in real-time to a computer for further processing. In addition, this sensor transmits the skin's temperature along with the hand's acceleration readings. Galvanic skin response can be used to detect user's arousal while interacting with the task or the robot. Data analysis will explore the reliability of this type of physiological data to measure user engagement.

3.1 OKAO evaluation

After it was decided to use OKAO as the main recognition SDK, we performed a number of tests in order to measure the performance and identify possible limitations of the system. We measured and evaluated the performance of OKAO in terms of smile detection using multiple cameras (three in total – see **Figure 1 left**) opposite to the user during a 45 second interaction with the screen and the robot. We used multiple cameras because the viewing angles significantly affect the performance of the facial recognition software. The user was instructed to interact with both the table and the robot whilst randomly smiling throughout the task varying the intensity.

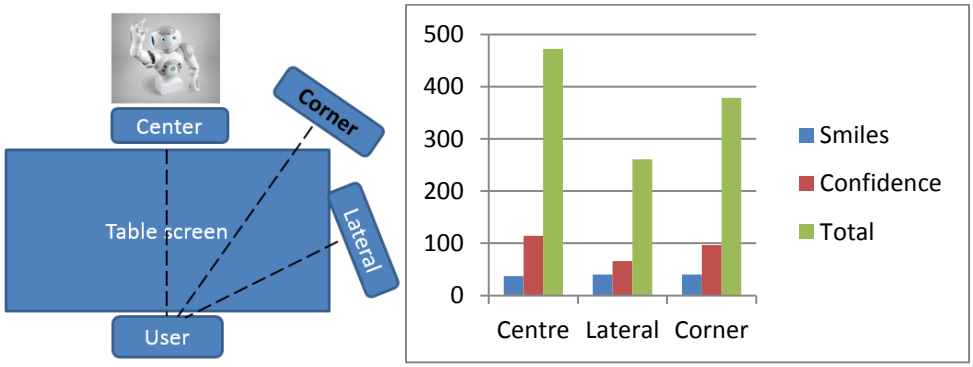

Figure 1: Camera setup and OKAO performance

The graph above (see **Figure 1 right**) shows the difference in terms of OKAO performance in detecting the average number of smiles of the user. The graph also presents the average

[6] Affectiva Q Sensor, www.affdex.com

confidence levels that were extracted from the software for each smile detection. In addition, we list the total values that have been calculated from the multiplication of gaze values and confidence levels. The calculated values elicit the average performance of smile using the confidence values as a complimentary measurement. The centre camera performs better than the other two as the user spends most of the time gazing at the screen thus, allowing the front camera to capture more facial information. The corner camera is a balance between the lateral and the centre in terms of performance. OKAO SDK is capable of recognizing the face even from a wide angle but a camera positioned at 45 degrees from the user boosts the confidence levels making it a considerably wiser choice than the lateral.

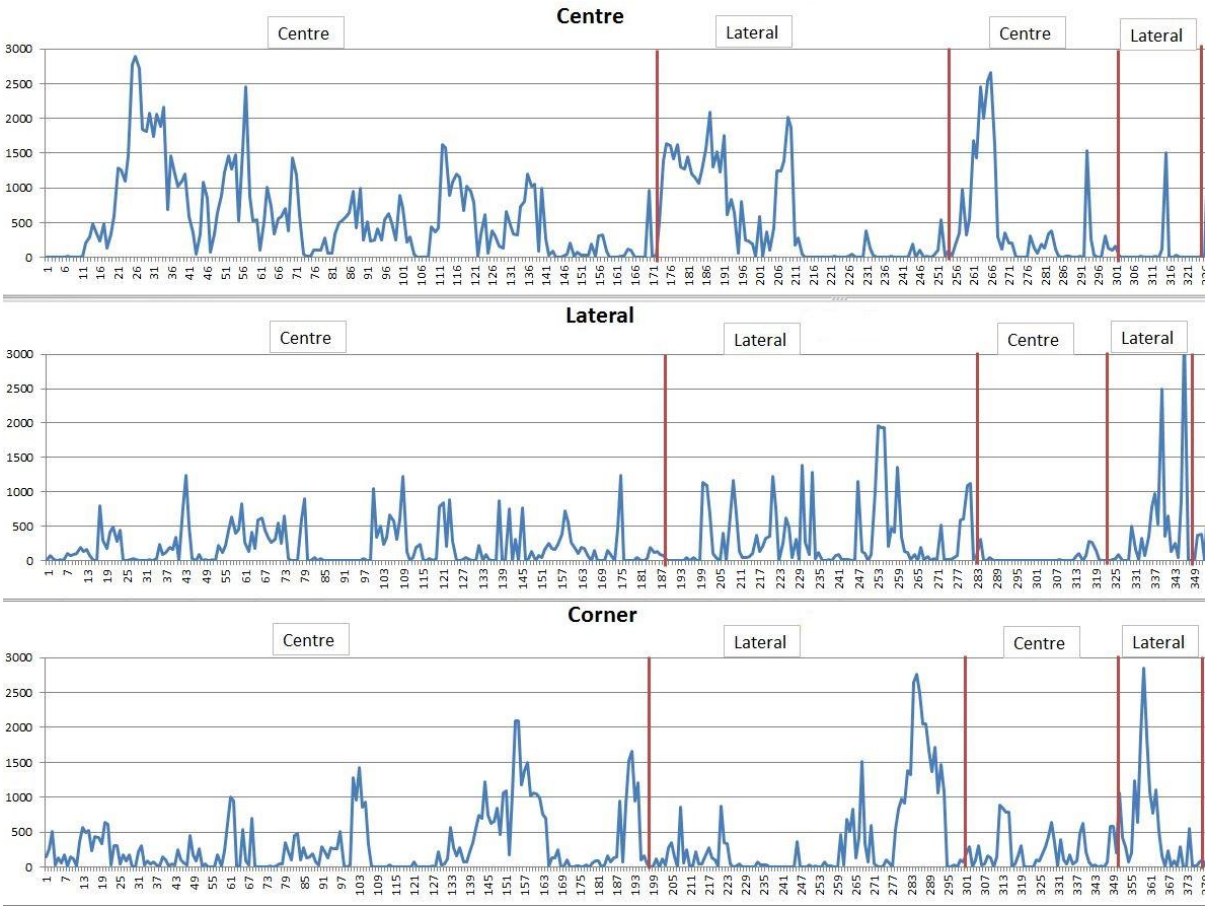

Figure 2: OKAO performance on smile detection

The graphs above (see **Figure 2**) represent the performance of OKAO during a 45 second interaction of a user with the screen and the robot. The graphs were drawn from calculating from input of both smile estimation and confidence levels in order to measure the performance of the system. The user kept shifting between facing an object (Robot) at the frontal and Lateral position, (Head orientation is shown in boxes above graphs) and kept smiling throughout the

task varying the intensity. Generally, OKAO recognizes the smile more accurately when the user is looking towards the active camera. Although OKAO still manages to recognize smile on extreme viewing angles, the confidence levels drop and false detections might occur.

4. User engagement pilot study

In WP4 we explore the development of a computational model capable of automatically detecting task and social engagement during human-robot interactions using streams of real-time sensory and contextual data. However, to begin, we must train the model to detect pre-defined indicators from within the streaming data. Therefore, we need to start by collecting a large corpus of multimodal data from real-time interactions involving a task and robot. From analysis of this corpus we can begin to explore reoccurring patterns to establish the most pertinent indicators of engagement.

The engagement pilot study is designed to collect a corpus which includes maximally opposite states of task-only engagement (i.e., from fully engaged to disengaged), and interactions with an engaging robot versus a less-engaging neutral robot during a collaborative task. The primary goal of the pilot study is to identify and explore possible indicators for user engagement with the task and the robot (in relation to Task 4.2). In order to identify these indicators a series of experiments were designed, developed and evaluated from a group of participants. These included a battery of short tests to extract engagement indicators such as lower and upper bounds for engagement detection within a task and social bonding between the participants and the robot.

Other objectives of these experiments are (1) to inform the design of a corpus collection via WoZ of the tutor's interaction with students in a classroom for training the affect recognition system to be developed in Task 4.3; and (2) to train and test initial prototypes of an automatic engagement detector (to be conducted as part of Task 4.3).

4.1 Set-up

The touchscreen's big dimensions (See Deliverable 6.1 for more information on hardware architecture) raised some questions regarding the position of the robot in relation to the user. In order to make the scenarios usable we came up with two possible setups: the frontal and lateral positions of the robot in relation to the user along with the different positions for the sensors. We sketched both conditions and positioned the sensors and cameras around the robot in order to prevent any possible obstruction in the camera view and maximize the

viewing angle as much as possible. The available sensors and the nature of the hardware allowed us to define two possible setups as shown in **Figure 3**.

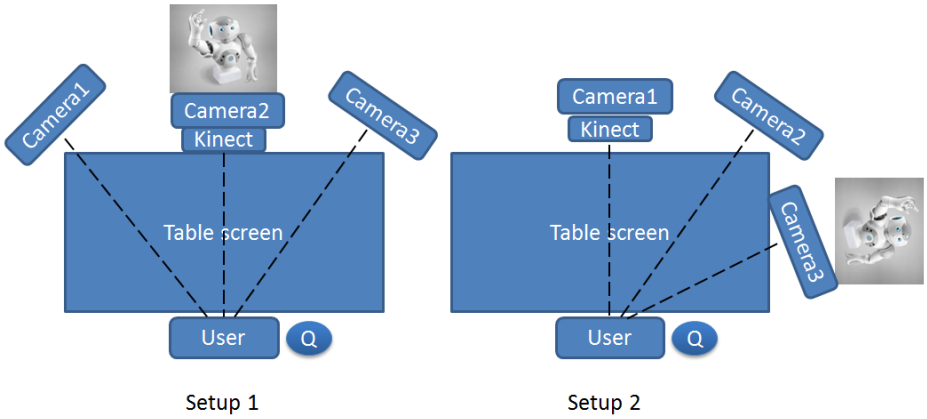

Figure 3: System setups

For the pilot study we decided to use and evaluate both possible setups in order to finalise our choice based on users' preference and statistical results. The evaluation and results of this study is further discussed in section 4.5 (Evaluation).

The technical implementation of our scenario (see **Figure 4**) is comprised of a large touch-screen table (A) to graphically represent the interactive tasks, a torso-only version of the NAO humanoid robot (B) to facilitate the social aspect of the interaction, several video cameras for detecting facial expressions (C), a Microsoft Kinect (D) for gaze direction and lean position relating to posture, and an Affectiva Q Sensor (E) for measuring galvanic skin response.

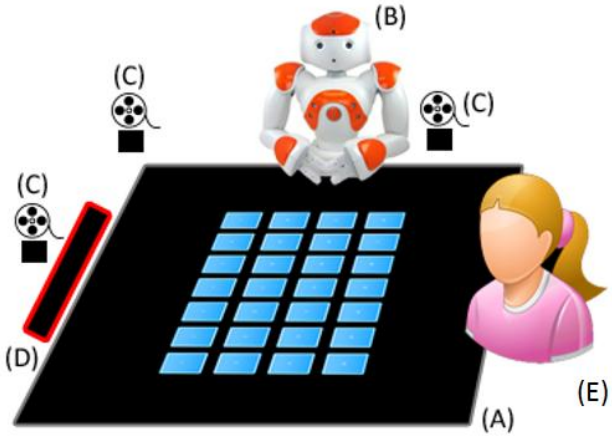

Figure 4 - Overview of the technical set up (Setup 2): (A) Touch Screen Table, (B) NAO Humanoid Robot (C) Cameras (D) Microsoft Kinect (E) Q Sensor

4.2 Experimental scenario

Our experimental scenario involves three interactive game-like tasks. The first two tasks are designed to elicit maximally opposite states of task engagement and the final task is designed to elicit social-task engagement. Together these tasks form the basis of two experiments and involved a total of 80 participants.

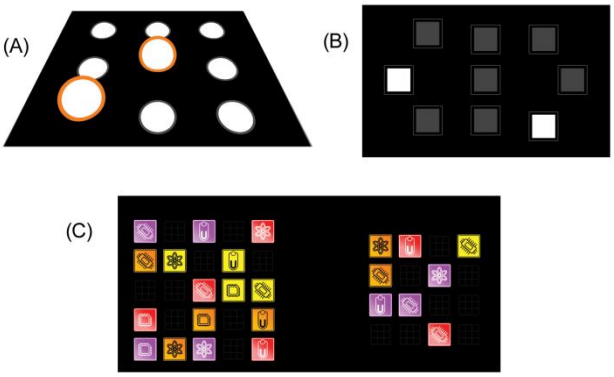

Figure 5 - Snapshots of (A) Task 1, (B) Task 2, and (C) Task 3.

Task 1

The first task is based on a simple 'Whack-A-Mole' style game and is considered to be an engaging task which requires effort and concentration (see **Figure 5 - A**). The participant interacts with this fast-paced game for three minutes by tapping on the screen when a "mole" pops up. This task uses both audio feedback and basic visual stimuli to engage the participant during play.

Task 2

The second task is based on the Corsi block-tapping test, and has been designed to be far less engaging than task 1. Participants are shown a sequence of three squares and they must repeat that exact sequence by tapping on the same squares in the same order (see **Figure 5 - B**). The sequences remain simple throughout the entire three minutes of the task. This task contains no audio beyond an initial countdown tone and very little visual stimuli, other than showing which block should be, or has been pressed.

Task 3

The third task is designed to observe social-task engagement during a novel human-robot interaction scenario. The task is a memory game involving cards and lasts for up to five minutes, although it can be completed sooner. The robot begins the interaction by explaining that his battery is damaged and he needs help to build a new one, thereafter, the human and the robot work together in an attempt to find all of the components. The participant is required to help the robot find the components it needs to build a new battery (see **Figure 5 - C**).

4.3 Methodology

Eighty-six adult participants were recruited from within the University of Birmingham. Participants were randomly divided into two main groups, representing the two overall conditions for the study (i.e., engaging and non-engaging). For an overview of all experimental conditions used in the pilot study see **Table 1**. A clearer explanation of how these conditions relate to each of the actual experiments is provided in the following sections.

Condition ID	Overall Conditions	Task Ordering	Robot Position	Participant Count
1	Engaging	1 - 2 – 3	Frontal	10
2	Engaging	1 - 2 – 3	Lateral	10
3	Engaging	2 - 1 – 3	Frontal	10
4	Engaging	2 - 1 – 3	Lateral	10
5	Non-Engaging	1 - 2 – 3	Frontal	10
6	Non-Engaging	1 - 2 – 3	Lateral	10
7	Non-Engaging	2 - 1 – 3	Frontal	10
8	Non-Engaging	2 - 1 – 3	Lateral	10

Table 1 - Experimental Conditions

4.3.1 Task Engagement (Experiment 1)

In the first experiment the robot was not present (see **Figure 6**) as we only wished to measure task engagement (Corrigan et al., 2013). Having participants undertake two different tasks would allow us to elicit two maximally different states of task engagement. We used a whack-a-mole style game to induce high levels of engagement (Task 1), and a far less engaging block-tapping task for inducing lower levels of engagement (Task 2). Here, participants from both main groups (i.e., engaging and non-engaging) were randomly divided into two sub-groups representing two further conditions. These conditions were: 1) play the engaging whack-a-mole style game followed by the less-engaging control task; and 2) undertake the less-engaging control task followed by the more engaging whack-a-mole style game. For clarification, this means that both sub-groups undertook both tasks, but one sub-group played the engaging task first whereas the other undertook the touch button control task first. This is standard practice to ensure the data we collect is not biased by the ordering of the tasks.

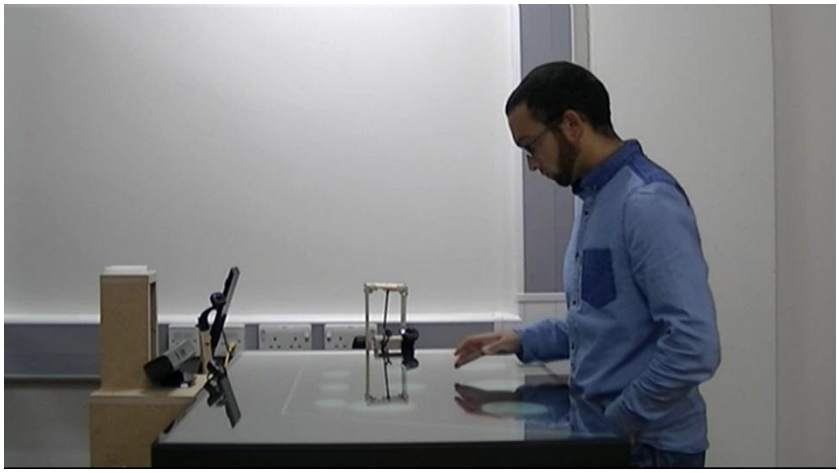

Figure 6: whack-a-mole style game.

4.3.2 Social-Task Engagement (Experiment 2)

In the second experiment participants were required to collaborate with the robot in a social task (Task 3). We had two primary conditions relating to the robots' behaviour and personality: 1) a helpful and instructive robot (deemed to be engaging), and 2) a neutral and partially-instructive robot (deemed to be non-engaging), and then two further conditions involving the position of the robot: 1) frontal, and 2) lateral. Therefore, participants from each of the overall conditions (i.e., engaging and non-engaging) were divided into two further groups representing these four conditions: 1) helpful and instructive robot in a frontal position (engaging group), 2) helpful and instructive robot in a lateral position (engaging group), 3) neutral and partially-instructive robot in a frontal position (non-engaging group), and 4) neutral and partially-instructive robot in a lateral position (non-engaging group). Here, we intend to analyse both social and task engagement between the different conditions. To prevent biasing the social relationship with the robot, participants had not been introduced to the robot prior to this experiment. **Figure 7** below shows the memory task and the robot.

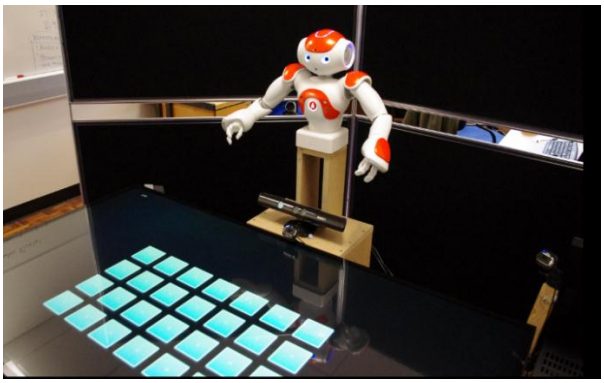

Figure 7: human-robot collaborative task.

4.3.3 Robot Behaviours

As discussed above, the second experiment consisted of two primary robot conditions relating to behaviour (i.e., helpful and instructive, and neutral and partially-instructive). These contrasting conditions allow us to explore the role that both social and task engagement played during real-world human-robot interactions.

Helpful and Partially Instructive Robot

Participants experienced a robot which was friendly, helpful and instructive, where the robots' behaviours are designed to be personable, pulling on the empathic strings of the participant. The robot described why 'they' need to work together to build 'their' battery, looking directly at the participant and addressing them by their first name. Here, the collaboration is emphasised and the participant is praised for performing well in the task.

Neutral and Partially Instructive Robot

In contrast, participants in this condition experienced a neutral and partially helpful robot which provided very little help and is far less personable. Here, the robot did not address the participant by name and does not explicitly refer to the collaboration using words such as "our" or "we" during the interaction. In terms of gaze, the robot would refrain from mutual gaze (i.e., eye to eye) in favour of a slightly downward gaze directed toward the table.

4.4 Data collection

The pilot study was performed in order to collect a corpus and interactions with an engaging robot/task versus a less-engaging neutral robot/task during a collaborative task. Therefore, the setup included a number of sensors in order to collect as much data as possible. Additionally, we handed questionnaires to the participants prior and after the study. Finally, the system was logging information from the task while the participants were interacting with the touchscreen.

4.4.1 Sensors

As described in Section 3, in order to capture and evaluate data from the study, we require a number of sensors capable of perceiving the environment and the users. For this study, we utilised three web cameras in different locations on top of the touch table, a Microsoft Kinect sensor, and an Affectiva Q Sensor. Since the purpose of the study was to collect data, we developed a simple client-server architecture in order to automate and synchronise the recording process. This architecture allowed us to develop individual modules for each sensor and distribute them in different computers. The module distribution helped with resource management as the sensors can run in real-time which consumes a significant amount of processing power. The three web cameras were pointing towards the participant and were saving compressed video files in a synchronised manner. The Q sensor was transmitting information to the main computer via Bluetooth with an 8 Hertz sample rate. Finally, the Kinect sensor extracted facial characteristics in real-time such as head direction information and facial action units along with the depth information.

The client-server architecture that we developed was linked with the main task which, in turn, controlled the sensors. Prior to each experiment, we inserted the user's ID and name and whenever the user selected to start the task, it sent a start signal to all of the connect sensor modules to start recording. At the end of each task, the main system sent a stop signal that triggered the saving method on each module.

Table 2 below displays the formatting of the log files:

Column No:	**Q Sensor Log**	**Kinect Log**
1	Date	Date
2	Time	Time
3	Packet number (ignore)	Participant's distance from sensor
4	Z Axis	Head Rotation X Axis
5	Y Axis	Head Rotation Y Axis
6	X Axis	AU26/AU27 Jaw Lower
7	Battery voltage	AU10 Lip Raiser
8	Skin temperature	AU13/15 Lip Corner Depressor
9	Eletrodermal activity (EDA)	AU20 Lip Stretcher
10		AU4 Brow Lower
11		AU2 Brow Raiser

Table 2: Log files

4.4.2 Questionnaire Design

Following each task, participants were asked to fill out a short survey questionnaire using a five-point Likert Scale (see **Appendix I**). For the task-only experiments participant questionnaires were related to task engagement using the Engagement Scale (see **Appendix I – Task Engagement 1-3**) (IRRE, 1998). The final task questionnaire was related to social engagement (see **Appendix I – Social Engagement Task 3**) and included two additional short personality-related scales from previous physiological research on empathy and loneliness, allowing us to test additional hypotheses concerning inter-individual differences and interactions with empathic beings (such as the robot). It is expected that there would be a significant correlation between empathic concern (of the participant) and social engagement.

4.4.3 Probes selection for engagement detection

Embedded Engagement Probes

A probe can be a non-intrusive, pervasive and embedded method of collecting informative data at different stages of an interaction (Corrigan et al., 2014). Here, the explicit probes (i.e., questionnaires) are embedded into the interaction itself, minimising disruption caused by

researcher intervention. The implicit probes are pervasive, executing at predefined stages within the interaction, minimising unnecessary disruption to the natural interactive flow. The feedback we gain from these probes is used as ground truths for engagement for training our models as well as milestones for other methods of data mining and statistical analysis.

A. Implicit Probes

1) Social Bonding Probe: This implicit probe occurs at the start of the third task, allowing the participant to choose whether or not they wish to interact with the robot.

2) Robot to Task: The second probe is also an implicit probe designed to measure the temporal lag involved with diverting gaze from the robot to objects which are graphically represented within the task. The results of this probe tells us how much attention the participant gives to the robot's implied instruction.

3) Task to Robot: During the interaction the robot will ask the participant a question which requires the participant to divert gaze and attention away from the task, towards the robot. Here, we detect any shift of gaze towards the robot and measure the initial temporal lag and further sustained gaze.

4) Attention to Instructions: If at any point during the task, the participant presses on a square whilst another is open, a buzzer will sound and the robot will inform the participant that they risk damaging the system if they press a square before the previous one has recovered. Following a warning, we measure the temporal difference of any future warnings.

B. Explicit Probes

1) Regular Self-Report Probes: The repeated self-report is embedded within the interaction to minimise disruption. This on-screen self-report probe occurs every minute, allowing us to measure changes of task and social engagement at different stages of the interaction.

4.5 Evaluation: initial results

Following the study, we performed an initial evaluation and analysis of the data, primarily to inform our pending work and planned studies. Our initial findings, based on the analysis of data collected from post experiment survey questionnaires and embedded explicit probes (i.e., on-screen self-report questionnaires) concluded that there are no significant differences in terms of engagement between the frontal and lateral robot positions. A non-parametric independent

samples test was used in SPSS[7] in order to evaluate the questionnaire data. Additionally, we statistically evaluated the data from the sensors in order to further investigate user's preference and performance while interacting with the task and the robot. This step has determined the final position of the main camera for facial recognition and furthermore, the most optimal robot position which is directly related with the sensors.

4.5.1 Robot position evaluation

In order to evaluate the optimal position for the robot, we divided the participants' data into 2 conditions, 42 participants with robot in lateral position and 43 in frontal position (see **Figure 1** for robot positions). For the Kinect data we measured the amount of times the sensor managed to successfully track users' face during their interactions. Furthermore, we calculated the average time (in percentage) the users spend facing towards the robot. Finally, we used the data derived from the distance sensor in order to calculate users' average distance from the table. For the Q Sensor we calculated the average values from the readings of electro-dermal activity and skin temperature. The results below display the comparison between the Kinect and Q Sensor analysis for the two proposed robot positions.

[7] SPSS, www.ibm.com/software/uk/analytics/spss/products/statistics

Frontal position

Kinect data

- Successful head tracking frames:83.68%
- Distance from table:1814.594
- User facing at robot:16.41%

Q Sensor data

- Electro-dermal activity:1.942681
- Skin temperature:34.48C

Lateral position

Kinect data

- Successful head tracking frames:80.98%
- Distance from table:1830.173
- User facing at robot:8.63%

Q Sensor data

- Electro-dermal activity:1.503776
- Skin temperature:34.16C

The results above are also supported by statistical analysis which revealed significant differences for both Kinect and Q Sensor data. An independent Samples test showed 0.0001 significance levels on the Kinect data (Depth: Sig.= 0.0001, t=-36.366, Gaze: Sig.= 0.0001, t=93.476) and 0.0001 on the Q Sensor data (Temp: Sig.= 0.0001, t=55.517, EDA: Sig.= 0.0001, t=58.347).

Recommended setup:

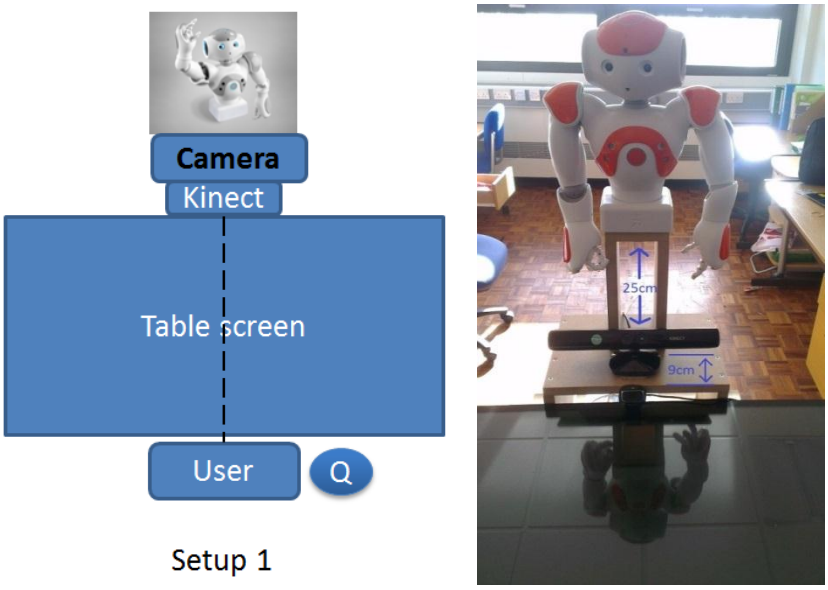

Figure 8: Recommended robot position and stand dimensions

Based on the analysis above, it was clear that the robot should be placed exactly at the centre of the long side of the table screen (see **Figure 8 left**), opposite to the user. A robot stand (see **Figure 8 right**) will support the robot 25cm higher than the table screen. The stand will increase robot's height and bring it closer to user's eye sight for potential mutual eye-gaze. Additionally, the stand will help with the location of the Kinect sensor and the frontal camera in a position not obstructed from robot's movements (hands).

In summary, placing the robot opposite the user will benefit the interaction and the engagement process as:

- Users spend more time facing towards the robot thus interacting with it more often
- Users tend to interact with the table in a closer proximity
- Electro-dermal activity is higher in this mode
- Skin temperature is higher

Further analysis of data collected in this pilot study will be conducted as part of Task 4.2 and described in deliverable D4.2.

5. WoZ study

5.1 Introduction

In order to build an intelligent artificial tutor with affective capabilities, an appropriate computational model needs to be developed and properly trained to automatically recognize and classify the emotional state of the user (i.e., here we focus on user engagement). For this reason, and based on the lessons learned from the pilot study on user engagement, a WoZ study was performed to collect multimodal data from school pupils aged between 10 and 13 interacting with the robotic tutor. In a WoZ study, a remote controlled agent is operated by a human wizard (a teaching expert in our WoZ study) in a controllable environment without the participant's knowledge that the robot is not autonomous. The Wizards' interface provides a level of control over the task and the robot. Here, we assume that the Wizard sits in place of the autonomous mind, giving the system the decisional ability of a human so we can learn more about how to facilitate the interaction between a child and the robotic tutor.

5.1.1 Role and control

As a very first step in the design of the interface we needed to decide exactly what role the Wizard plays in the system and how much control he/she should have via the interface. Should the wizard be a controller, a moderator or a supervisor (Dow et al., 2005)?

Controller

If the wizard is a Controller, then he/she is concerned with even the lowest level control of the robots' behaviour. This is relatively simple to design as we can simply categorise each of the behaviours and lay them out so that the Wizard chooses each and every behaviour needed during the interaction. However, the problem we may face with this level of control is that the Wizard will become consumed with remembering and choosing the right behaviours and will pay less attention to what is actually happening within the task and during the interaction. With the Controller role, we would expect a certain temporal lag in-between actions performed by the user in the task and the Wizard being able to choose the correct behaviour, speech act and gaze direction, in addition to any computational and mechanical lag derived from the equipment.

Moderator

At a mid-level, a Moderator is more concerned with ensuring that the robot's behaviour is suitable for the actions performed by the user in the task. Here, the robot's behaviours are predefined, but still not fully automated. The Wizard is presented with a sub-set of behaviours and only has to choose one from that sub-set to moderate the behaviour selection. Using a behavioural hypotheses for both the user and robot during each stage of the task (Green et al., 2004), we can highlight the behaviours which suit that particular stage of the task or actions selected by the user. The wizard is then able to select behaviours from the highlighted sub-set, whilst still being able to select something else if they wish to. This method would reduce some of the temporal lag experienced with the Controller.

Supervisor

At an even higher level, the Wizards' role is to supervise the interaction. Here, the Wizard is abstracted away from the common low-level behavioural selection completely, but remains in control from a higher level. The wizard does still have several low-level controls for providing the user with feedback during task stages and for commenting on several basic unrelated task elements, yet common behavioural selection is automatically selected from a set based on several other higher level factors, such as the amount of help and scaffolding the Wizard feels is appropriate (i.e., based on the pedagogical strategies). Here, the Wizard can concentrate on

making the interaction more about the learning content where the system is closest to that of the final EMOTE vision.

With the Supervisor, there is more time to facilitate the interaction and concentrate on what the user is doing and how to best support the user within the task, not on what should be said or gestured when answers are right or wrong (i.e., the Controller role). The interface can be focused more towards the pedagogical strategies that one should employ during the different stages of the task.

5.2 Methodology

The WoZ study was conducted in a classroom in the Arthur Terry School, a secondary school in Birmingham, United Kingdom. Hardware for both the Wizard and the touch screen were set up in the same room. The reason for that was the wizard wanted to have direct sound feedback from the participant without any delays that might have caused from wireless transmission. Additionally, having the wizard and the participants in one room would allow us to monitor both in case of an error or synchronisation issues. The touch screen and the related hardware were placed in one side of the room and we used some tall dividers in order to separate the room into two sections. The wizard was seated in the other corner of the room hidden behind the dividers and the monitors.

Before each session, the participant was instructed to fill a short questionnaire while we put a Q Sensor on their wrist in order to warm up and provide accurate readings. After the questionnaire, the participant was walked to the touch screen and was given instruction about the task. The wizard started the study and controlled the task progression throughout the session providing help when needed via the robot. During each session, we manually captured three videos via digital camcorders. Two camcorders were placed around the touch table to record participants' interactions and the running task and the third was placed behind the wizard in order to capture the control panel screen along with the wizard's comments. Additionally, we positioned three web cameras around the touch screen in order to provide the wizard a clear picture of the task and the participant interacting with it. The duration of each session varied between 10 and 20 minutes as the wizard controlled the steps and the progression. At the end of each session, the participant was instructed to fill another questionnaire about the task. If time allowed, the wizard explained to the participant that the robot was not autonomous but controlled from the wizard and at the same time interviewed the participant about the task and their overall experience.

The list below displays the required technical material to run the WoZ study.

- 3x web HD cameras
- 1x Kinect sensor version 1
- 1x Q Sensor
- Multi-Touch table
- 3x Camcorders
- 3x Tripods
- Screens to separate the area
- 1x NAO Robot
- 1x Server for sensor logging
- 1x computer for WoZ
- 1x network router

5.2.1 Participants

For the WoZ study we recruited 20 participants from the Arthur Terry School in Birmingham. 9 participants were female while 11 were males and they were all in School year 7 (11-13 years old).

5.2.2 Learning scenario

The WoZ is built upon the findings from the pilot study and was conducted in a real classroom environment with learning content having been informed and supported by teaching experts. Additionally, teaching experts informed and supported the content of the learning task with the purpose of increasing its pedagogical value and reflect the children's needs and requirements. In this section, we provide a description of the learning scenario used in the WoZ study. Work conducted in WP2 (see **Deliverable 2.1** for details) provided a set of requirements for the implementation of the learning scenario.

The learning scenario focuses on local map reading as an individual activity. The scenario is easily adaptable to local maps and allows for adaptation of difficulty levels. The scenario involves local map reading skills based around a back-story that consists of following a trail to find the best location to place a new statue in an art trail.

The student has to find several informants on the map for which the robotic tutor gives directions in terms of map reading competencies (e.g., "go South 200 meters"). Difficulty levels can be varied by a) giving more complex directions, b) using more difficult versions of map reading skills (e.g., using South-South-West instead of just South or requiring to calculate), or c)

using more difficult concepts (e.g., altitude instead of compass directions). For each informant that is found, the student receives a clue (or several clues) about the hiding place of the person/object. The clues are also defined in terms of map reading competencies (e.g., the person/object is hiding in a woody area). The same way of adapting difficulty levels as for the first step can be used in the clues. After finding all informants, and thereby receiving all clues, the student has to point out the place of the person/object from a set of possible locations. Only one location fulfils all constraints provided by the clues, all other locations are incorrect in at least one of the constraints, thereby making it possible to detect flaws in the student's thinking. This phase requires the student to combine clues, which is inherently a more difficult task. Using more or fewer possible locations is a way to adapt the difficulty level of this task.

5.3 System architecture

The system was developed using a robotic embodiment as the tutor, and a large touch table for the interaction. The touch table runs a multimedia interactive application (like a videogame). The user interacts with the system by using the game application. The system also interacts back through the robotic embodiment, which provides a character with expressive behaviour. **Figure 9** shows how the current system is structured for the WoZ experiment. Our physical components are the NAO robot, a multi-touch table (MTT), and a Microsoft Kinect.

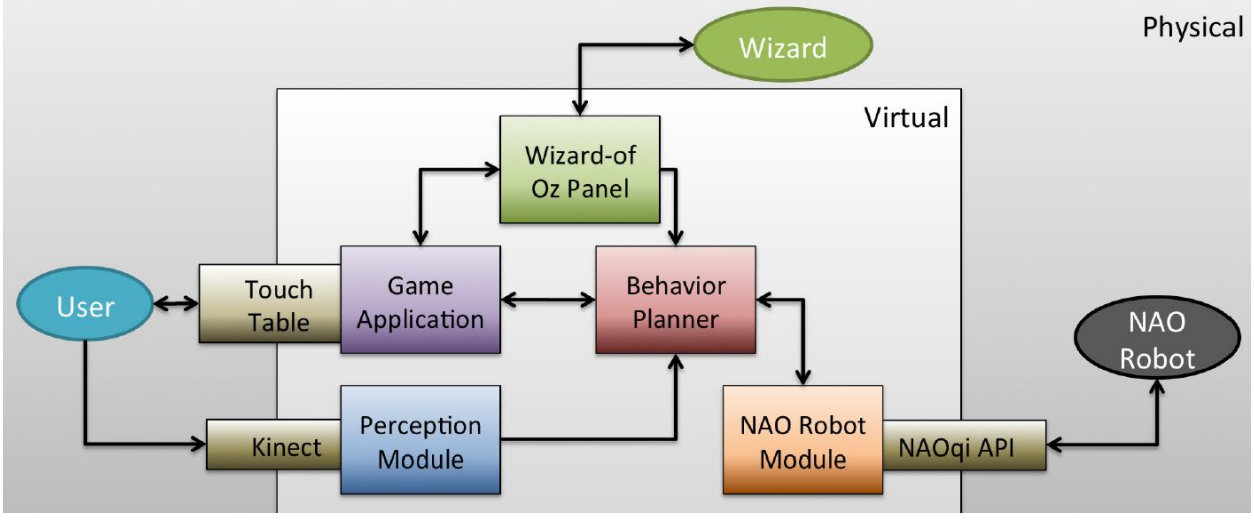

Figure 9: System for WoZ

The MTT provides both a virtual environment (Game Application) that is shared by the agent and the user, and is also used for input from the user. The Kinect captures the user, being currently used only for head tracking. The NAO Robot provides an embodiment that exhibits expressive behaviour towards the user. Such expressive behaviour is generated and managed by a Behaviour Planner (BP) module. This module is further described in its dedicated section. The Wizard uses the WoZ panel to control the flow of the game, to parameterise some of the BP's semiautonomous behaviour, and to manually select high level FML utterances. These utterances are dialogue acts which were previously written and tagged both with nonverbal and game instructions. The FML is broken down in the BP into BML actions and game actions. The actions are then sent through Thalamus to be scheduled and/or routed to other modules (NAO Robot Module and Game Application). **Figure 10** shows the overall system architecture for the WoZ study.

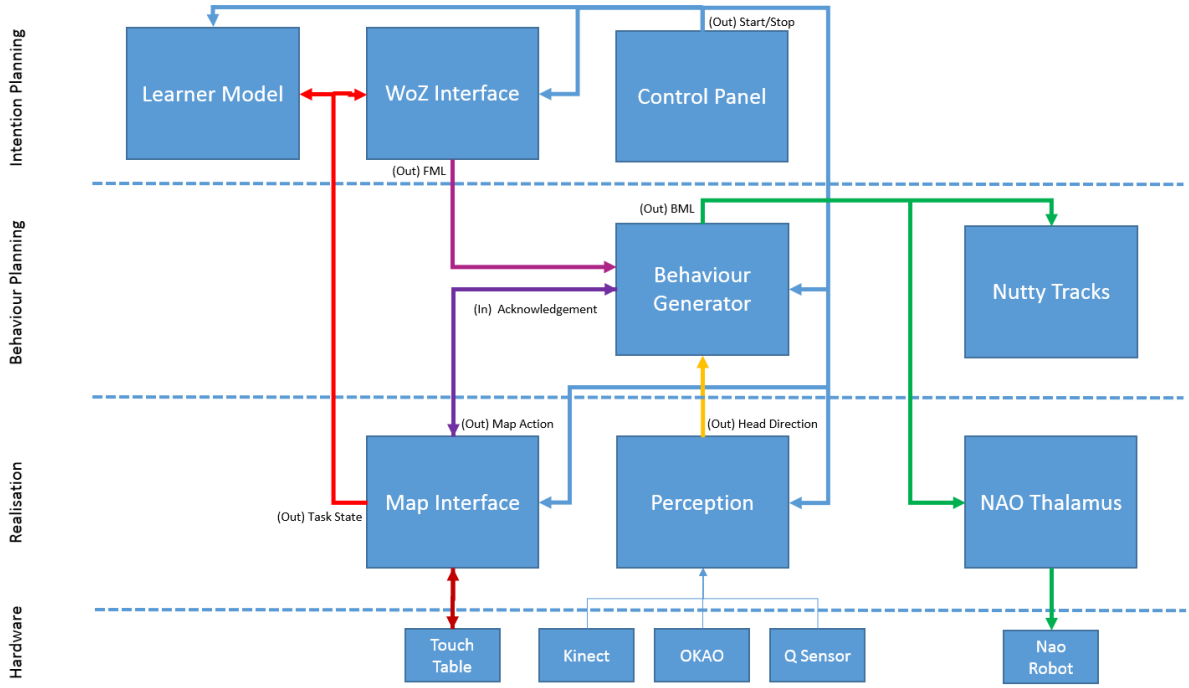

Figure 10: System architecture for WoZ

5.3.1 Implementation for WoZ

Given the above, an implementation has been made for a WoZ study (see **Figure 11**). We created a script and interaction design document to detail at each step in the task what could happen.

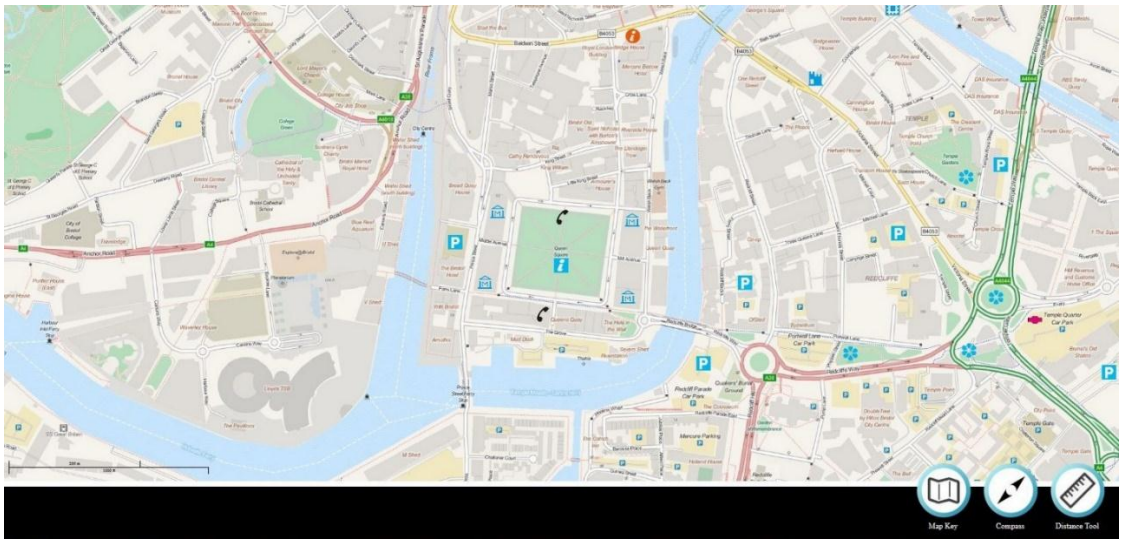

Figure 11: Interface for the map and tools

5.3.1.1 Interaction design

Based on the requirements derived as part of the work done in WP2, it is important that the tutor can comment on how well the student is doing and also gesture and look at elements on the screen, and in some cases interact.

For this to work the system communicates the following to the rest of the system:

- Location of objects on screen
- Where and what the learner has clicked on the screen or map
- If the learner has answered the question correctly or incorrectly and a breakdown of this, as it is possible to get answer partially correct. This information is essential for the rest of the system to make decisions.

The task can also listen for messages from the rest of the system:

- Allow the tutor to control elements of the task.
- Highlight tools and areas of interest on the map.

Static information is shared by a task script.

5.3.1.2 Technology implementation

After completing a technical review of possible technologies to implement the learning scenario, we chose a web app to display and interact with a map. The app is built using OpenLayers Toolkit [8](an open source Java-script library for displaying and interacting with map data), with additional controls implemented using HTML and Java-script. The web app back end is written in Java, which handles communication with the rest of the system via Thalamus. Map data for

[8] Openlayers Toolkit, http://openlayers.org/

the activity will be stored and served locally using GeoServer [9] (also open source). A combination of OpenLayers Toolkit and GeoServer effectively creates a robust application which does not require an Internet connection to run (a key requirement for teachers).

We have also developed an authoring tool so that teachers and even learners can create trails and activities within the scenario.

5.3.2 Control Panel for the Wizard

Once we had established that the wizard would need to have control as both the moderator and supervisory roles, we could look into ways of designing controls that are intuitive and easy to use. This was an iterative process using different design methodologies, where the Wizard was involved in each stage of the design.

Figure 12: Control panel

Interface controls for the different aspects of the system (i.e., camera control, system control, competency grading, task script, and utterance and feedback selection) were separated according to know interface design principles (see **Figure 12**) (Benyon, Turner, & Turner, 2005).

[9] GeoServer, http://geoserver.org

Camera Control

The interface can display live feed from the three cameras involved in the system. The frontal and lateral camera views of the participant are housed side-by-side in 320x240 containers in the top left hand corner of the interface, and below is a larger 640x480 container displaying the top-down view of the interactive screen with the participants' hands clearly shown interacting with the task. This larger area and resolution allows the Wizard to see the task in a fairly high level of detail.

System Control

Positioned in the lower left-hand corner is a small control panel, which allows the Wizard to start and stop the system (in terms of video recording, data logging, and task progression). Here, the Wizard can enter the participants' details (i.e., name and ID), which is then fed into the system so data logs can be saved along with the specific participant ID and the robot can address the participant by name.

Competency Grading Control

The interactive map task scenario performs some basic competency grading during the interaction. The architecture allows the interface to read in this information and to display it on-screen along with additional controls, which allow the Wizard to specify their own grading. Later analysis can be used to train the scenario's grading system.

Task Script Control

The predefined task script is loaded directly into the interface upon initialisation. Displaying the different steps of the task script along with a clear indicator of the participants' current position within the script allows the Wizard to stay abreast of the task as it unfolds. Additional controls allow the Wizard to repeat the question, retry the step and progress onto the next step.

Utterance and Feedback Selection Control

Predefined pedagogical strategies and feedback utterances are loaded directly into the interface upon initialisation. Here, the Wizard has full control over which utterance is selected and sent onto the behaviour planner for processing. In their raw format utterances contain the actual speech act along with animation and gaze/glance tags. For fast viewing and selection by the Wizard these tags are removed from the utterance before being displayed on-screen.

Tools Control

The participant can use tools to help them with the task (i.e., compass, measuring tool, map key). This control allows the Wizard to show and hide the tools from this interface.

5.3.3 Robot Behaviours

In order to perform the WoZ, behaviours for the Nao robot were created. The creation of these behaviours resembled on literature review and inspiration from human studies performed within the scope of the project (see **D2.1**). The development of these behaviours is described below.

Verbal behaviour

Translations and transcriptions

In order to have a better organisation of the material, the first step was to transcribe and translate the video recordings from Swedish to English from mock-up 2 held in Sweden.

Coding scheme

Having data transcribed and translated, a literature review regarding appropriate coding schemes for ITS and ATS (e.g., Graesser, Wiemer-Hastings, Wiemer-Hastings, & Kreuz, 1999) was performed in order to understand the main dimensions of verbal behaviour that human teachers perform during the learning process. Therefore initial conceptual categories of scaffolding strategies were considered according to literature and inspired by the interaction within the specific task developed for mock-up 2 (see **D2.1**) (e.g., pump, prompt, hint, etc.). These scaffolding strategies categories were rearranged until the verbal behaviour of this context could be included. The creation of extra verbal categories was addressed as a way to complement the verbal dimensions found in previous literature and provide a natural course of the interaction flow (e.g., small talk, stalling, and guidance).

The categories were also combined with the specific task script developed by UoB for the WoZ geography task for an individual learning scenario. Therefore, step-related utterances related to each task step were formulated in a similar way to the way teachers deliver these utterances.

In addition, utterances that would resemble those conveyed by a peer or learning companion relating to the appearance of the robot were created. The reasoning behind this is that teachers are capable of a performance in a shared physical space that robots are unable to (e.g. facial

expressions). As such, we tried to overcome Nao's limitations by providing expressive features that specifically would fit into Nao's embodiment.

Thereafter, the final categories for verbal behaviour consist of 25 different strategies related to the different learning skills of the task and the context-dependent dynamics required (see **D5.2**).

Nonverbal behaviour

Nonverbal behaviour (e.g., gestures) is essential in all types of communication and was therefore considered for the educational context of a robotic tutor in the WoZ study. The nonverbal behaviour was mainly inspired by the video recordings from the first and second mock-up studies (see **D2.1**) and literature on the theme (e.g., McNeill, 1992). The intention is to provide expressiveness to the robot enabling the student to engage in a natural educational interaction. The nonverbal behaviour considered different types of expressiveness through animations (e.g., emotion expression (happy)) and gestures (e.g., deictic gestures signalling where the student should direct his/her attention) (see **D5.2**).

Behaviour of the robot

The verbal and nonverbal behaviours were adapted to Nao's embodiment and capacities, enabling us to create an artificial tutor based on human studies in a pre-scripted behaviour generation for the Wizard. Teacher mimicry is considered very important for the various scaffolding strategies, and whereas it is obvious that teachers do not perform all be behaviours that the robot did (e.g., teachers do not raise their arms in excitement when delivering positive feedback), we had to overcome the limitations of the robot and adapt human behaviours to an artificial agent.

The work developed by INESC-ID & UGOT provided all the verbal and nonverbal behaviours for the WoZ study. It is also important to note that the continuous process of defining behaviours also attended to a very active and close participation of the wizard along the process.

Apart from this, inspiration from human studies and literature regarding ITS and ATS was performed. The combination of verbal and nonverbal behaviours to increase task engagement and learning outcomes and robot's expressiveness was also addressed, as well as the adaptation to the robot embodiment.

5.3.4 Behaviour Planner

The tutor's behaviour is all managed by the Behaviour Planner (BP) module. This module provides a link between semiautonomous and WoZ triggered behaviour. It is constantly updated with information from the Perception module (which interfaces with the Kinect). It also receives coordinate information from the Game Application, in order to be able to instruct the robot to gaze, point or wave towards specific points on the screen. The BP also provides some semiautonomous behaviour. It manages gazing behaviour so that the Wizard does not have to deal with all that. The semi-automated gazing behaviour is currently implemented such that the BP is instructed on which the current Gaze Target is for the Tutor (the student, a specific screen point, or wherever the user is clicking). This target can be selected manually by the Wizard, or can be included in the utterances that the Wizard selects to be performed. The gazing targets, however, do not keep the tutor gazing towards a single direction. If the target is the student, then the robotic tutor will gaze track the student, while also performing gaze aversion towards a random direction. In order to have the tutor follow on the user's task performance, the gaze target is set to the Clicks, and as such, the BP will keep the tutor gazing at wherever the student clicks. If the student takes some time to click, the tutor will quickly glance towards the student once in a while. The BP also schedules and performs multimodal utterances that are selected by the Wizard. These multimodal utterances are tagged with instructions about animations, gazing, pointing and even game behaviours that the tutor should execute while performing a speech. By coupling all the nonverbal behaviour with the utterances library, it made it easier to enrich the nonverbal performance of the tutor while not burdening the Wizard with a lot of controls.

5.4 Data collection

Simultaneous video feeds from the cameras, the Q sensor, live OKAO and the Kinect sensor were recorded during the tutor-learner interaction. The interaction between the tutor and the learner in terms of tutor dialogue actions, utterances and learner responses in terms of button presses were also be logged.

5.4.1 Sensors

This study utilises the same capturing equipment as in the pilot study but differently developed as the server now is Thalamus. We developed Thalamus modules for each sensor, one for the web camera, one for Kinect, one for Q Sensor and one central called Perception manager. **Figure 13** below represents the position for each sensor in relation to the user.

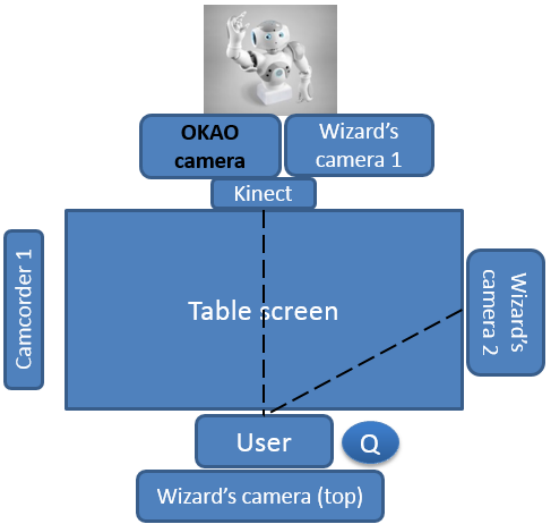

Figure 13: Sensors and robot position

The web camera module uses the OKAO libraries in order to extract real-time facial characteristics such as expression estimation and smile estimation, eye gaze information and blink estimation. This information is being transmitted in real-time back to the perception module which in turn saves or uses the timeframes. The web camera is located under the robot and focuses on participant's face.

The Kinect module uses the Microsoft Kinect sensor in real-time to extract head gaze information, depth and facial action units. All of this information is being transmitted back to the Perception manager in real-time via pipes.

The Q Sensor module handles the serial communication between the device and the computer and transmits the electro-dermal activity and temperature back to the perception module.

The Perception module handles the communication between the sensor modules and the rest of the system and receives the messages from the sensors. Depending on the set sample rate, it stores the incoming data in chunks for easier evaluation. Additionally, this module uses a neural network for estimating the gaze direction of the user in real-time based on three variables from the Kinect sensor: head pan, head tilt and head height. The neural network has been trained using three inputs, six branches and one output. The training was performed using 3 participants with different heights while interacting with the touch table and performing various gazing. An observer classified their gaze into a database which was later used to train

the network. The network had four types of outputs: Gaze at robot, Gaze at left part of screen, Gaze at right part of screen and Gaze elsewhere.

5.4.2 Questionnaire Design

In the development of the post-experimental questionnaires for the WoZ we focused on important features of task and social engagement, empathy levels, and features of attachment and socio-emotional bonding. The use of paper version questionnaires was the best format for gathering this qualitative and quantitative data; it is a simple format for children, at the age of 11-13 to understand and use. Paper questionnaires typically allow researchers to gather large quantities of data in a short period of time with limited effects on validity or reliability (Patton, 2005). For use within the WoZ, it also allowed us to quickly and easily quantify the data collected as all ratings had numeric values. Therefore, after quantifying the data, we were able to compare and contrast the results, from different participant conditions in the WoZ.

The first questionnaire (see **Appendix II – Task Engagement – Social Engagement**) examined task and social engagement. Task engagement (6 items in questionnaire) is related specifically to how engaging a task is perceived to be, both affectively and cognitively (McGregor & Elliot, 2002). The questionnaire was given to participants after the task was completed and it intended to measure the participant's immediate reaction to the task and based on this reaction, their motivation and action level in regards to the specific task (McGregor & Elliot, 2002). Social engagement (10 items in questionnaire) refers to how individuals interact with others socially, the level to which we want to participate in collective activities and how we form and maintain social relationships and connections with others (Ryan & Patrick, 2001). Social engagement also requires the interaction between a minimum of two parties; in the EMOTE context it will involve the interaction between child and robot, or children and robot. The questions related to social engagement were written to gauge the enjoyment, depth, breadth and level of interaction between the child and the robot. In addition questions also asked about the motivation to either assist the robot as an interaction partner and/or the desire to continue socially engaging with the robot.

The second questionnaire (see **Appendix II – Empathic concern scale**) was used to examine and understand levels of empathy in students aged 11-13. Results from the empathy questionnaire will be used to evaluate whether empathy levels impact task and social engagement and socio-emotional bonding between student and robot. We are looking to assess whether there are differences in empathic capabilities that will affect the interaction with the artificial agent. The development of varied subscales that measure trait-like differences are mainly tested on adult

populations (Chlopan et al., 1985). For use in the WoZ, we used a pre-existing empathy measure which examines levels of dispositional empathy in children and youth, Bryant's Empathy Index (BEI), (Bryant, 1982). The BEI examines empathy from a multidimensional approach and includes subscales similar to that of the Interpersonal Reactivity Index (IRI), (Davis, 1983). The BEI and IRI have been tested for reliability and validity (Zhou, Valiente, & Eisenberg, 2003). The subscales of the BEI include: 'feelings of sadness', 'understanding feelings', and 'tearful rejection' (Bryant, 1982). In a validation study, it has been found that the BEI provides a very comprehensive, multidimensional measure of empathy, examining both the cognitive and affective components of empathy in a child-friendly/comprehensive manner (Aristu, Tello, Ortiz & Gandara, 2008).

The third questionnaire (see **Appendix II – Attachment Scale**) (8 items) explored the topic of attachment by creating questions to examine socio-emotional bonding. These questions were adapted from the scales developed by Schifferstein and Zwartkruis-Pelgrim (2008). In the EMOTE context we want to develop questions that will assess the empathic connection and socio-emotional bond between the student artificial agent (robotic tutor). This third questionnaire was designed to assess perceived levels of closeness, attachment, meaning and strength of feelings between the child and the robot. This post-experiment questionnaire examines how close the child relates to the robot and its impact level. This questionnaire can be used after one or multiple interactions with the robot. The level of attachment or socio-emotional connection is hypothesized to be positively correlated to amount of time spent interacting with the robot, based on research on interaction and attachment (Ryan & Patrick, 2001.

5.4.3 Probes selection for engagement detection

Probes allow us to collect event-driven temporally-specific feedback directly from the user without interrupting the interaction. For engagement, collecting data discretely in this way means we can gather reliable context-specific data from within the interaction itself. Furthermore, we can concentrate our post-experiment analysis within short windows of time when noteworthy events actually took place. Overall, this is an important concept, allowing us to specify which element of the system the participants are mostly engaged with at specific times. Probes also allow us to circumvent problems which may otherwise arise from typical post-experimental questionnaires, such as bias from 1) subsequent events to the one we are interested in, 2) the overall interaction, and 3) other behavioural aspects such as embarrassment or excitement related to task progress.

For the WoZ study, we decided that only implicit engagement probes would be used. This was based upon advice from a teaching professional that our explicit probes (i.e., on-screen self-report questionnaires) could interrupt the children mid-flow. Furthermore, the implicit probes were selected from within the pre-designed interaction script as oppose to retro-fitting the probes into the interaction post-design. Doing so removed some flexibility in their construction, but helped to ensure that each step of the interaction was context-specific and natural in respect of flow, rather than having additional steps which appear pointless or out of place. Ultimately, these probes will allow us to automate the process of labelling segments of continuous data, ideal for building a computational model capable of automatically detecting the learner's engagement from live data feeds.

5.4.3.1 Social Bonding Probe

At the beginning of the interaction the robot introduces himself as NAO and asks the child for their name. Here, we will measure 1) if the child states there name in response, and 2) if so, was the child gazing at the robot at the time. Then, after a short pause (i.e., allowing time for the child to answer) the robot states that it is nice to meet the child. Here, we will measure whether the child smiles in response. In addition to this, throughout this entire probe we are also looking to measure how much gaze is directed toward the robot during this introductory period, whilst also allowing for short glances elsewhere.

5.4.3.2 Robot to Task Probe

Almost immediately after the robot has introduced himself, he informs the child that there are tools at the bottom right hand side of the screen. Here we will measure 1) whether or not there was a shift in gaze from the robot toward those tools, and 2) if so, is there any temporal difference between cases which might correlate with engagement level derived from post-experiment domain-specialist coding and video annotation. This probe will help us to understand whether the child is solely engaged with the novelty of the robot's presence, or whether they are also engaging with the robot as an interactive partner.

5.4.3.3 Task to Robot Probe

Throughout the task the robot will often provide guidance in terms of new information which is relevant to the question, or to provide feedback which is designed to encourage or discourage certain actions being taken. Using this probe we intend to measure 1) whether the child breaks gaze from the task toward to the robot, or 2) whether the child maintains gaze with the task but adapts their actions to suit the robots guidance, or 3) whether the child ignores the robots guidance completely. This probe will help us to understand whether the child is engaged solely with completion of the task at hand, or whether they are engaged with both task and robot enough to self-regulate the interaction.

5.4.3.4 Attention to Instruction

The interaction is designed with several navigational map trails, each consisting of several steps. The child must be able to respond to each of those steps appropriately (by interaction on the touch screen) in order to progress through the trails. When the child makes a mistake or fails to understand what is expected of them, the robot will intervene and repeat or reiterate the important information which is contained within that step. Here we will concentrate on whether the child was able to correct their response having been given this new information, and how many times the information needed to be given before a correct response was possible.

6. Conclusions and future work

Task 4.1 was concerned with a corpus collection via a WoZ study of the tutor's interaction with students in a classroom for training an affect recognition system developed in Task 4.3. We presented implementation and initial results of a pilot study which aimed to evaluate system setup, sensors and probes for automatic engagement detection. The pilot study informed the best robot position and the most optimal setup for the sensors. Further analysis of data collected during the pilot study will be performed and described in Task 4.2. Based upon lessons learnt by conducting the pilot study, UoB, supported by the other partners involved in WP4, conducted a WoZ experiment with early secondary students in order to collect learner-tutor interactions (video recordings of students' behavioural expressions and synchronised contextual information) in the Scenario 1 developed in WP2. The WoZ study provided us with a large amount of data that will be analysed in Task 4.2.

In summary, the objectives of Task 4.2 are:

- Identification of indicators conveying successful information for learner affect prediction performed in Task 4.3
- Design and implementation of behavioural and contextual indicators for learner affect prediction
- Design and implementation of indicators of the learning progress based on the learner's abilities and difficulties measured through the learner's actions on the learning platform

Task 4.2 aims to establish the foundations for the tutor's perception of the learner. The design of an affect recognition system and a learner model requires knowledge of what behavioural and contextual indicators are informative for a specific recognition or modelling task and when, during the interaction, the tutor requires information about how the learner is feeling or

progressing with the learning task. Results from the pilot experiments on engagement described in Task 4.1 will inform the design of a framework for the automatic recognition of the learner's engagement. Data analysis is planned for the second half of year 2 aiming to identify the most relevant behavioural and contextual indicators of engagement and include them in the development of the automatic affect detector planned in Task 4.3.

Another objective of Task 4.2 is to identify and develop indicators of the learner's progress measured through the learner's actions on the learning platform. Based on the mock-up studies on Scenario 1 conducted in WP2 in collaboration with UGOT, UoB designed and implemented a skeleton version of the learner model. This includes a representation of the student in the system, particularly in terms of learner competencies relating to the scenario (i.e., map reading skills such as distance, compass direction, and map symbol), history of questions (e.g., results and time taken to complete exercise), and history of usage of tools/hints. In year 2 this representation of the learner in the system will be taken further, and a learner model built following the investigation of techniques such as constraint-based modelling.

The outputs of this task will be described in deliverable D4.2, due in M24.

References

- Aristu, A. L., Tello, F. P. H., Ortiz, M. Á. C., & Gándara, M. V. D. B. (2008). The structure of Bryant's Empathy Index for children: A cross-validation study. *The Spanish journal of psychology*, *11*, 670-677.
- Awais, M., & Henrich, D. (2010). Human-robot collaboration by intention recognition using probabilistic state machines. In *Robotics in Alpe-Adria-Danube Region (RAAD), 2010 IEEE 19th International Workshop on*, (pp. 75 –80).
- Benyon D, Turner P, Turner S. (2005) *Designing interactive systems*. Harlow, England: Addison-Wesley
- Bryant, B. K. (1982). An index of empathy for children and adolescents. *Child Development*, 53, 413-425.
- Castellano G., Leite I., Pereira A., Martinho C., Paiva A., and McOwan P. W., (2012) "Detecting engagement in HRI: An exploration of social and task-based context," in *2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, pp. 421–428.
- Chlopan, B. E., McCain, M. L., Carbonell, J. L., & Hagen, R. L. (1985). Empathy: Review of available measures. *Journal of personality and social psychology*, 48, 635-653.
- Corrigan, L. J., Peters, C., & Castellano, G. (2013). Identifying Task Engagement: Towards Personalised Interactions with Educational Robots. *In Affective Computing and Intelligent Interaction (ACII),* 2013 Humaine Association Conference on (pp. 655-658). IEEE.
- Corrigan L.J., Basedow, C., Küster, D., Kappas, A., Peters, C., and Castellano, G. (2014). Mixing implicit and explicit probes: finding a ground truth for engagement in social human-robot interactions. *In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI '14)*. http://doi.acm.org/10.1145/2559636.2559815
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.
- Dow et al, (2005). "Wizard of Oz support throughout an iterative design process," Pervasive Computing, IEEE , vol.4, no.4, pp.18,26
- Graesser, A. C., Wiemer-Hastings, K., Wiemer-Hastings, P., & Kreuz, R. (1999). AutoTutor: A simulation of a human tutor. *Cognitive Systems Research,1(1),* 35-51.
- Glowinski, D., & Mancini, M. (2011). Towards real-time affect detection based on sample entropy analysis of expressive gesture. *In Proceedings of the 4th international conference on Affective computing and intelligent interaction - Volume Part I*, ACII'11, (pp. 527–537). Berlin, Heidelberg: Springer-Verlag.
- Green, A., H¨uttenrauch, H., & Eklundh, K. S. (2004). Applying the Wizard-of-Oz framework to cooperative service discovery and configuration. *In 13th IEEE International Symposium on Robot and Human Interactive Communication (Ro-MAN)* (pp. 575–580). IEEE. http://dx.doi.org/10.1109/ROMAN.2004.1374824
- Gunes, H., & Schuller, B. (2012). Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image and Vision Computing.*
- IRRE (1998) Research Assessment Package For Schools (RAPS) Manual, Philadelphia. Pa: Institute for Research and Reform in Education
- Jung, H., Ou, S., & Grupen, R. (2010). Learning to perceive human intention and assist in a situated context. *RSS workshop on Learning for Human-Robot Interaction Modelling.*

- McGregor, H. A., & Elliot, A. J. (2002). Achievement goals as predictors of achievement-relevant processes prior to task engagement. *Journal of Educational Psychology, 94, 381-395*. doi: 10.1037//0022-0663.94.2.381
- McNeill, D. (1992). Hand and mind: What gestures reveal about thought. University of Chicago Press.
- Patton, M. Q. (2005). Qualitative research. John Wiley & Sons, Ltd.
- Poggi I. (2007) Mind, hands, face and body. A goal and belief view of multimodal communication. Weidler, Berlin.
- Schifferstein, H. N., & Zwartkruis-Pelgrim, E. P. (2008). Consumer-product attachment: Measurement and design implications*. International Journal of Design, 2, 1-13*.
- Sidner C. L., Kidd C. D., Lee C., and Lesh N., (2004) "Where to look: A study of human-robot engagement," *in Proceedings of the 9th international conference on Intelligent user interfaces.* ACM, pp. 78–84.
- Rich C., Ponsleur B., Holroyd A., and Sidner C. L.,(2010) "Recognizing engagement in human-robot interaction*," in 5th ACM/IEEE International Conference on Human-Robot Interaction, HRI 2010*, pp. 375–382.
- Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal, 38*, 437-460. doi: 10.3102/00028312038002437
- Shic, F., Chawarska, K., Bradshaw, J., & Scassellati, B. (2008). Autism, eye-tracking, entropy. *In Development and Learning, 2008. ICDL 2008. 7th IEEE International Conference on, (pp. 73–78)*. IEEE.
- Sakita, K., Ogawara, K., Murakami, S., Kawamura, K., & Ikeuchi, K. (2004). Flexible cooperation between human and robot by interpreting human intention from gaze information. *In Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on, vol. 1, (pp. 846 – 851 vol.1).*
- Shernoff D. J., Csikszentmihalyi M., Schneider B., and Shernoff E. S., (2003) "Student engagement in high school classrooms from the perspective of flow theory," *School Psychology Quarterly, vol. 18, no. 2, pp. 158–176.*
- Shindev, I., Sun, Y., Coovert, M., Pavlova, J., & Lee, T. (2012). Exploration of intention expression for robots. *In Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, (pp. 247 –248).
- Zhou Q., Valiente C., & Eisenberg N. (2003). Empathy and its measurement. In S. J. Lopez (Ed.) and C. R. Snyder (Ed.), *Positive psychological assessment: A handbook of models and measures, 269-284*. Washington, DC: American Psychological Association.

Appendix I: Questionnaire for engagement study

Participant Questionnaires

.....

Task Engagement (Task 1)

.....

	Not at all			Neutral			***Very Much***
I enjoyed this task.	1	2	3	4	5	6	7
I found this task challenging.	1	2	3	4	5	6	7
I would like to continue to do this task.	1	2	3	4	5	6	7
It was important for me to do well on this task.	1	2	3	4	5	6	7
I found this task easy to understand.	1	2	3	4	5	6	7
I found this task interesting.	1	2	3	4	5	6	7
I have previous experience with similar tasks.	1	2	3	4	5	6	7

.....

Task Engagement (Task 2)

.....

	Not at all			Neutral			***Very Much***
I enjoyed this task.	1	2	3	4	5	6	7
I found this task challenging.	1	2	3	4	5	6	7
I would like to continue to do this task.	1	2	3	4	5	6	7
It was important for me to do well on this task.	1	2	3	4	5	6	7
I found this task easy to understand.	1	2	3	4	5	6	7
I found this task interesting.	1	2	3	4	5	6	7
I have previous experience with similar tasks.	1	2	3	4	5	6	7

.....

Task Engagement (Task 3)

.....

	Not at all			Neutral			***Very Much***
I enjoyed this task.	1	2	3	4	5	6	7
I found this task challenging.	1	2	3	4	5	6	7
I would like to continue to do this task.	1	2	3	4	5	6	7
It was important for me to do well on this task.	1	2	3	4	5	6	7
I found this task easy to understand.	1	2	3	4	5	6	7
I found this task interesting.	1	2	3	4	5	6	7
I have previous experience with similar tasks.	1	2	3	4	5	6	7

.....

Social Engagement (Task 3)

.....

	Not at all			Neutral			***Very Much***
I would like to play another game with the robot.	1	2	3	4	5	6	7
I feel the robot was like a partner in the task.	1	2	3	4	5	6	7
I believe the robot cared about whether I did well in the task.	1	2	3	4	5	6	7
I was worried about the robot when the battery started to run out.	1	2	3	4	5	6	7
I would have felt badly if I would not have been able to fill up the robots battery.	1	2	3	4	5	6	7
I wanted to keep helping the robot.	1	2	3	4	5	6	7

I felt like the robot and I were part of the same team.	①	②	③	④	⑤	⑥	⑦
I found the robot considerate.	①	②	③	④	⑤	⑥	⑦
I found the robot helpful.	①	②	③	④	⑤	⑥	⑦
I found the robot friendly.	①	②	③	④	⑤	⑥	⑦

Empathic Concern Scale

	Not at all			Neutral			*Very Much*
I often have tender, concerned feelings for people less fortunate than me.	1	2	3	4	5	6	7
Other peoples misfortunes do not usually disturb me a great deal.	1	2	3	4	5	6	7
When I see someone being taken advantage of, I feel kind of protective toward them.	1	2	3	4	5	6	7
When I see someone being treated unfairly, I sometimes don't feel very much pity for them.	1	2	3	4	5	6	7
I would describe myself as a pretty soft-hearted person.	1	2	3	4	5	6	7
Sometimes I don't feel sorry for other people when they are having problems.	1	2	3	4	5	6	7
I am often quite touched by things that I see happen.	1	2	3	4	5	6	7

UCLA

	Not at all			Neutral			*All the Time*
How often do you feel that there is no one you can turn to?	1	2	3	4	5	6	7
How often do you feel alone?	1	2	3	4	5	6	7
How often do you feel part of a group of friends.	1	2	3	4	5	6	7
How often do you feel close to people?	1	2	3	4	5	6	7
How often do you feel that your relationships with others are not meaningful?	1	2	3	4	5	6	7
How often do you feel isolated from others?	1	2	3	4	5	6	7

How often do you feel that there are people who really understand you?	① ② ③ ④ ⑤ ⑥ ⑦
How often do you feel that people are around you but not with you?	① ② ③ ④ ⑤ ⑥ ⑦
How often do you feel that there are people you can talk to?	① ② ③ ④ ⑤ ⑥ ⑦
How often do you feel that there are people you can turn to?	① ② ③ ④ ⑤ ⑥ ⑦

Appendix II: Questionnaire for WoZ study

Participant Questionnaire

Please note that there are no right or wrong answers in this questionnaire. The following questions are not concerned with your abilities, but how you perceived the robot and the game. At the end, there are also a few questions about how you see yourself and others, and how you deal with feelings and emotions.

ID:	
Gender:	**Male / Female**
Experience with Computers:	① ② ③ ④ ⑤
Experience with Map Reading:	① ② ③ ④ ⑤
Experience with Robots:	① ② ③ ④ ⑤

.....

Task Engagement

.....

		Not at all Neutral ***Very*** ***Much***
1	I enjoyed this activity	① ② ③ ④ ⑤
2	I found this activity hard	① ② ③ ④ ⑤
3	I would like to continue with this activity	① ② ③ ④ ⑤
4	It was important for me to do a good job	① ② ③ ④ ⑤
5	I found this activity easy to understand	① ② ③ ④ ⑤
6	I have done activities like this before	① ② ③ ④ ⑤
7	I enjoyed this activity	① ② ③ ④ ⑤
8	I did well in this activity	① ② ③ ④ ⑤
9	I did my best in this activity	① ② ③ ④ ⑤

.....

Social Engagement

.....

		Not at all	Neutral	*Very Much*
1	I would like to play another game with the robot	1	2 3 4	5
2	I feel the robot was like a partner in this activity	1	2 3 4	5
3	I believe the robot cared about whether I did well in the activity	1	2 3 4	5
4	I would be worried about the robot if the battery started to run out	1	2 3 4	5
5	I wanted to keep helping the robot	1	2 3 4	5
6	I felt like the robot and I were part of the same team	1	2 3 4	5
7	I found the robot caring	1	2 3 4	5

.....

Attachment Scale

.....

		Not at all	Neutral	*Very Much*
1	I feel very close to the robot	1	2 3 4	5
2	This robot is special to me	1	2 3 4	5
3	I have a bond with this robot	1	2 3 4	5
4	This robot has no special meaning for me	1	2 3 4	5
5	I am very attached to this robot	1	2 3 4	5
6	This robot has a special place in my life	1	2 3 4	5
7	This robot means a lot to me	1	2 3 4	5
8	I have no feelings for the robot	1	2 3 4	5

Empathic Concern Scale

1	It makes me sad to see a girl who can't find anyone to play with.	**True / False**
2	People who kiss and hug in public are silly.	**True / False**
3	Boys who cry because they are happy are silly.	**True / False**
4	I really like to watch people open presents, even when I don't get a present myself.	**True / False**
5	Seeing a boy who is crying makes me feel like crying.	**True / False**
6	I get upset when I see a girl being hurt.	**True / False**
7	Even when I don't know why someone is laughing, I laugh too.	**True / False**
8	Sometimes I cry when I watch T.V.	**True / False**
9	Girls who cry because they are happy are silly.	**True / False**
10	It's hard for me to see why someone else gets upset.	**True / False**
11	I get upset when I see an animal being hurt.	**True / False**
12	It makes me sad to see a boy who can't find anyone to play with.	**True / False**
13	Some songs make me so sad I feel like crying.	**True / False**
14	I get upset when I see a boy being hurt.	**True / False**
15	Grown-ups sometimes cry even when they have nothing to be sad about.	**True / False**
16	It's silly to treat dogs and cats as though they have feelings like people.	**True / False**
17	I get mad when I see a classmate pretending to need help from the teacher all the time.	**True / False**
18	Kids who have no friends probably don't want any.	**True / False**
19	Seeing a girl who is crying makes me feel like crying.	**True / False**
20	I think it is funny that some people cry during a sad movie or while reading a sad book.	**True / False**
21	I am able to eat all my cookies even when I see someone looking at me wanting one.	**True / False**
22	I don't feel upset when I see a classmate being punished by a teacher for not obeying school rules.	**True / False**